DOCUMENT RESUME

ED 311 076 TM 013 916

AUTHOR	Kulick, Edward; Hu, P. Gillian
TITLE	Examining the Relationship between Differential Item Functioning and Item Difficulty. College Board Report No. 89-5.
INSTITUTION	College Entrance Examination Board, New York, N.Y.; Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-89-18
PUB DATE	89
NOTE	37p.
AVAILABLE FROM	College Board Publications, Box 886, New York, NY 10101-0886 ($7.00).
PUB TYPE	Statistical Data (110) -- Reports - Research/Technical (143)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Asian Americans; Black Students; College Bound Students; *College Entrance Examinations; Comparative Testing; *Difficulty Level; Ethnicity; Higher Education; High Schools; Hispanic Americans; Mathematics Tests; *Racial Differences; Reading Tests; *Sex Differences; Test Bias; *Test Items; Verbal Tests; White Students
IDENTIFIERS	*Differential Item Performance; Mantel Haenszel Procedure; *Scholastic Aptitude Test

ABSTRACT

The relationship of differential item functioning (DIF) to item difficulty on the Scholastic Aptitude Test (SAT) was examined, based on data from nine recent administrations of the test from June 1986 through December 1987. This pool of information includes item statistics on 765 verbal and 540 mathematical items computed for subgroups of White, Hispanic, Black, Asian American, male, and female examinees. In general, item difficulty is related to DIF. The nature of the relationship appears to be independent of the choice of DIF indices (either the Mantel-Haenszel or the standardization approach). The relationship was not dependent on test form. However, the relationship was dependent on the particular group comparison and on both the test sections and the item type being analyzed. The relationship was strong for each of the racial and ethnic group contrasts (in which Black, Hispanic, and Asian American examinees were compared in turn with White examinees), but was weak for the female and male examinee contrast. The relationship also appeared stronger on the verbal than on the mathematical sections. The relationship is such that more difficult items tended to exhibit positive DIF (the DIF favored the focal group over the White reference group). On the verbal sections, only the reading comprehension item type failed to exhibit a strong relationship. An index of the relative difference in omit rates between groups matched in ability correlated very highly (negatively) with DIF. Although DIF and this latter index are mathematically dependent measures, it was proposed that the latter index may be partly responsible for the relationship between DIF and item difficulty. Hispanic and Black focal groups tended to omit differentially less than did the White reference groups; for Asian Americans the opposite held. Eight figures and 17 data tables are provided. (TJH)

TM

College Board Report

No. 89-5

311076

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

ROBERT G. CAMERON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Examining the Relationship between Differential Item Functioning and Item Difficulty

Edward Kulick
P. Gillian Hu

Examining the Relationship between Differential Item Functioning and Item Difficulty

Edward Kulick
P. Gillian Hu

College Board Report No. 89-5
ETS RR No. 89-18

College Entrance Examination Board, New York, 1989

Edward Kulick is a lead research data analyst at Educational Testing Service, Princeton, New Jersey.

P. Gillian Hu is an associate measurement statistician at Educational Testing Service, Princeton, New Jersey.

Acknowledgments

The authors would like to thank Neil J. Dorans and P. W. Holland of ETS for their comments and reviews of earlier versions of this report.
Funding for this report was provided by the College Board/ETS Joint Staff Research and Development Committee.

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a nonprofit membership organization committed to maintaining academic standards and broadening access to higher education. Its more than 2,600 members include colleges and universities, secondary schools, university and school systems, and education associations and agencies. Representatives of the members elect the Board of Trustees and serve on committees and councils that advise the College Board on the guidance and placement, testing and assessment, and financial aid services it provides to students and educational institutions.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101-0886. The price is $7.

Copyright © 1989 by College Entrance Examination Board. All rights reserved.

College Board, Scholastic Aptitude Test, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board. Printed in the United States of America.

CONTENTS

Abstract 1

Introduction 1

Content Description of the SAT 3

Method 3

Results: SAT-Verbal Sections 5
- Hispanics and Whites 5
- Blacks and Whites 6
- Asian Americans and Whites 6
- Females and Males 7

Results: SAT-Mathematical Sections 7
- Hispanics and Whites 7
- Blacks and Whites 7
- Asian Americans and Whites 7
- Females and Males 7

Discussion 8

Summary 9

References 10

Figures

1. Mantel-Haenszel Delta-Difference versus Equated Delta (Hispanics and Whites) 11

2. Mantel-Haenszel Delta-Difference versus Differential Percentage Omitting (Hispanics and Whites) 12

3. Mantel-Haenszel Delta-Difference versus Equated Delta (Blacks and Whites) 13

4. Mantel-Haenszel Delta-Difference versus Differential Percentage Omitting (Blacks and Whites) 14

5. Mantel-Haenszel Delta-Difference versus Equated Delta (Asian Americans and Whites) 15

6. Mantel-Haenszel Delta-Difference versus Differential Percentage Omitting (Asian Americans and Whites) 16

7. Mantel-Haenszel Delta-Difference versus Equated Delta (Females and Males) 17

8. Mantel-Haenszel Delta-Difference versus Differential Percentage Omitting (Females and Males) 18

Tables

1. Regression Results for Predicting MHD from SAT-Verbal Item Characteristics (Hispanics and Whites) 19

2. Correlations Based on SAT-Verbal Data for Hispanic Focal Group and White Reference Group 20

3. Means and Standard Deviations of Mantel-Haenszel Delta Difference Index (MHD) 21

4. Regression Results for Predicting MHD from SAT-Verbal Item Characteristics (Blacks and Whites) 22

5. Correlations Based on SAT-Verbal Data for Black Focal Group and White Reference Group 23

6. Regression Results for Predicting MHD from SAT-Verbal Item Characteristics (Asian Americans and Whites) 24

7. Correlations Based on SAT-Verbal Data for Asian American Focal Group and White Reference Group 25

8. Regression Results for Predicting MHD from SAT-Verbal Item Characteristics (Females and Males) 26

9. Correlations Based on SAT-Verbal Data for Female Focal Group and Male Reference Group 27

10. Regression Results for Predicting MHD from SAT-Mathematical Item Characteristics (Hispanics and Whites) 28

11. Correlations Based on SAT-Mathematical Data for Hispanic Focal Group and White Reference Group 28

12. Regression Results for Predicting MHD from SAT-Mathematical Item Characteristics (Blacks and Whites) 29

13. Correlations Based on SAT-Mathematical Data for Black Focal Group and White Reference Group 29

14. Regression Results for Predicting MHD from SAT-Mathematical Item Characteristics (Asian Americans and Whites) 30

15. Correlations Based on SAT-Mathematical Data for Asian American Focal Group and White Reference Group 30

16. Regression Results for Predicting MHD from SAT-Mathematical Item Characteristics (Females and Males) 31

17. Correlations Based on SAT-Mathematical Data for Female Focal Group and Male Reference Group 31

ABSTRACT

This study examined the relationship of differential item functioning (DIF) to item difficulty on the Scholastic Aptitude Test (SAT). The data comprise verbal and mathematical item statistics from nine recent administrations of the SAT. In general, item difficulty is related to DIF. The nature of that relationship appears to be independent of the choice of DIF index (either the Mantel-Haenszel or the standardization approach) as well as of test form. However, the relationship was dependent on the particular group comparison and on both the test sections and the item type being analyzed.

The relationship was strong for each of the racial and ethnic group contrasts—in which black, Hispanic, and Asian American examinees were compared in turn with white examinees—but was weak for the female and male examinee contrast. The relationship also appeared stronger on the verbal sections than on the mathematical sections. The relationship is such that more difficult items tended to exhibit positive DIF (DIF favored the focal group over the white reference group). On the verbal sections, only the reading comprehension item type (with the smallest observed range in item difficulty) failed to exhibit a strong relationship.

Another index, the standardized difference in percentage omit (DIFPOM), correlated very highly (negatively) with DIF. Differential omission refers to a relative difference in omit rates between groups matched in ability. In fact, DIFPOM was consistently a better predictor of DIF in most models than was item difficulty. The relationship between DIF and DIFPOM held up across all four comparisons, including gender. It was also present in the mathematical sections with nearly the same magnitude exhibited in the verbal sections.

Although DIF and DIFPOM are mathematically dependent measures, it was proposed that DIFPOM may be partly responsible for the relationship between DIF and item difficulty. To what extent DIF is a consequence of differential omission and to what extent differential omission is a manifestation of DIF is problematic. Nonetheless, the presence of differential omission on a test has the *potential* to influence DIF indices and therefore should be an important concern especially for formula-scored tests, where omission occurs often on difficult items.

Among other findings is that Hispanic and black focal groups tended to omit differentially less than did the white reference groups. For Asian American examinees, the reverse holds. For females and males, the direction depends on the test sections. In general, groups that omitted differentially less experienced a relative advantage (high-positive DIF values) on the more difficult items, as measured by the DIF indices studied here (which treat omits as wrong in their calculation).

INTRODUCTION

Differential item functioning (DIF) has become an important issue in ability measurement and test fairness. It is defined as differential performance between two groups on an item after the groups have been matched with respect to the ability or trait that the item is purported to measure. For a test that measures a single dominant trait and for examinees at the same ability level, DIF refers to the phenomenon that a test item is more difficult for examinees in one of the two groups being compared. Differential item functioning is an item characteristic that provides important information about major subgroups of the test-taking population. In other words, DIF assesses the interaction of item characteristics (more difficult versus less difficult) with group membership (reference group versus focal group). The study of DIF not only benefits the test development specialist but also enhances the educator's understanding of various subgroups in terms of their cognitive processes, test-taking strategies, and knowledge deficiencies. The purpose of this study is to investigate DIF in more depth from the aspects of item characteristics and examinee response patterns. To be specific, the relationship between DIF and another important item characteristic, item difficulty, is explored. In addition, the predictive ability of differential omission rates of the groups for DIF is also examined.

Over the years items of the Scholastic Aptitude Test (SAT) have been analyzed for DIF at Educational Testing Service (ETS). The documentation of these analyses indicates a consistent finding for black examinees when they were compared with matched white examinees; i.e., the analogy items of the SAT-verbal test were differentially more difficult for the blacks (Dorans 1982; Kulick 1984; Rogers, Dorans, and Schmitt 1986; Rogers and Kulick 1987). The results of five studies conducted between 1975 and 1979, which examined black and white candidate performance on items from the SAT and the Test of Standard Written English (TSWE), were summarized and synthesized by Dorans (1982). A salient finding was summarized in this review. For two of the studies that employed both the delta-plot method and a log-linear method to detect DIF, all analogy items with extreme DIF were found to favor whites. The standardization approach (Dorans and Kulick 1983) also showed on different data that the analogy items tended to favor whites (Kulick 1984; Rogers, Dorans, and Schmitt 1986). Rogers and Kulick's (1987) study confirmed by the standardization method that the analogy item type exhibited more DIF than did other item types for blacks. Given these empirical results for blacks, this study aims at exploring the relationship between DIF and item difficulty on SAT analogy items for blacks and other minority groups, including Hispanics, Asian Americans, and females.

Factors that may be related to DIF have been identified for black examinees on the analogy items by Schmitt and Bleistein (1987). The list of related factors includes differential speededness, position within analogy item set, subject-matter content, and vertical relationships, among others. There are 10 analogy items within each SAT-verbal section. Since the items are ordered by difficulty, from easiest to hardest, Schmitt and Bleistein divided each analogy item set into the first five items and the last five items. The results obtained by the standardization method showed that the first five items account for most of the negative DIF found for blacks. In other words, blacks were found to perform worse on easier analogy items than on more difficult ones. An exact item difficulty measure was not used by Schmitt and Bleistein because such a measure is interdependent with other factors (e.g., subject-matter content and level of abstractness) they studied in the paper. In the current study, an overall item difficulty measure is used with the understanding that it comprises effects of the many factors that contribute to item difficulty.

Using item position as a measure of item difficulty for SAT analogy items, Freedle and Kostin (1987) found item difficulty to be a significant predictor of DIF for blacks. Items that appear earlier within the analogy set (easier items) yield negative DIF values, and items that appear later within the set (more difficult items) yield positive DIF values. Similar findings were reported by Freedle and Kostin (1988) for the Graduate Record Examinations (GRE). Item position and actual rank difficulty (based on the percentage correct of each item) were each used as a measure of item difficulty. The GRE-verbal test has four item types: antonyms, analogies, sentence completion items, and reading comprehension items. Evaluation of the relationship of DIF with item difficulty (measured by the actual rank difficulty) showed stronger correlations for analogy items and antonym items than for the other two item types. Analogy items and antonym items usually appear in very limited verbal contexts in comparison with sentence completion and reading comprehension items. An explanation offered by the authors states that "it is the augmenting and diminution of multiplicity of meanings that may be operating differentially across these four item types" (Freedle and Kostin 1988, 33). Similar findings were also reported by Freedle and Kostin in their paper for the SAT-verbal test, which has the same item types as the GRE-verbal test. Given these findings, the present study investigates the relationship between DIF and item difficulty for all item types for the SAT.

Besides a correct answer, possible responses to a multiple-choice item include omitting the item, not reaching the item, and choosing a distractor. Studying these response patterns helps us define the nature of the DIF associated with a particular item. The current SAT instructions consist of guidelines about both guessing and omitting an item under the formula-scoring direction (*Taking the SAT*, 1988). Examinees are encouraged to guess from the remaining options if they know one or more options are definitely wrong. The examinees are informed that they may omit items and that they neither gain nor lose credit by doing so. Given these instructions, it is of interest to study the relationship of DIF to the omit patterns of the groups being compared. In particular, the fact that blacks performed better, relative to comparable whites, on more difficult analogy items than on easier ones may partially be attributable to their differences in omit patterns, especially since the DIF measures used operationally at ETS are obtained by scoring items right or wrong regardless of scoring instructions.

The standardization procedure lends itself to the evaluation of examinee response patterns (Dorans and Kulick 1986). After the groups are matched on the total test score, the standardized percentage difference between the groups may be calculated for all other responses by replacing correct responses with responses of omits, not-reached items, and distractors (Dorans, Schmitt, and Bleistein 1988; Rivera and Schmitt 1988).

In their study of the SAT analogy items, Schmitt and Bleistein (1987) found differential omit patterns for whites and blacks. Whites tended to omit more on the items that seem differentially more difficult for them. Instead of omitting the items that appear differentially more difficult for them, blacks tended to choose a distractor, either through guessing or by the use of other strategies (e.g., vertical relationships). Freedle and Kostin (1988) concluded that differential omit rates between whites and blacks do not have a significant effect on DIF for the GRE analogy items. However, they suggested some evidence of more substantial differences in omission rates for the SAT analogy items. Rivera and Schmitt (1988) studied omit patterns for whites and Hispanics on the SAT and concluded that Hispanics generally omit less than do whites of comparable ability. Based on these findings, this study explores the predictive power of differential omit patterns on DIF for various minority groups. Since there is some indication of interaction between item type and omit patterns in Rivera and Schmitt's study, the relationship between DIF and omit patterns is also examined in that context.

Results from previous research indicate that easier, rather than harder, analogy items are differentially more difficult for blacks, relative to comparable whites. The relationship between DIF and item difficulty is explored on SAT analogy items for black examinees, as is the generalizability of this finding to other minority groups including females, Hispanics, and Asian Americans. The generalizability of this finding to other SAT-verbal item types and to SAT-mathematical item types

is also examined for all groups. Blacks and Hispanics were found to omit differentially less relative to whites in previous research (Schmitt and Bleistein 1987; Rivera and Schmitt 1988). The present study also explores the relationship between DIF and differential omit patterns for all gender and ethnic comparisons in an effort to help explain the relationship found between DIF and item difficulty. Other factors that are explored include item discrimination, test form effect, and the interaction of item difficulty with both item type and test forms.

CONTENT DESCRIPTION[1] OF THE *SAT*

Each SAT test book is divided into six separately timed 30-minute sections: two SAT-verbal sections (a total of 85 questions); two SAT-mathematical sections (a total of 60 questions); one TSWE section (50 questions); and one variable section, which does not count toward students' scores.

The 85 questions in the two verbal sections of the SAT are made up of four types: antonyms, 25 questions; analogies, 20 questions; sentence completion items, 15 questions; and reading comprehension items, 25 questions. Antonym questions are used to test breadth and depth of vocabulary. Analogies test a student's ability to establish relationships between pairs of words and to recognize similar or parallel relationships in other pairs.

Sentence completion questions test a student's ability to recognize logical relationships among parts of a sentence. Sentences are given in which one or two words have been omitted. The correct answer is the word or set of words that, when inserted in the blanks, best fits the meaning of the sentence as a whole. Reading comprehension questions are based on reading selections that have been adapted from published materials to make them suitable for testing. The selections vary in length (typically between 200 and 450 words) and in content. Reading questions test comprehension at several levels. Some questions ask the student to recognize a restatement of specific information contained in the passage; others ask the student to recognize main ideas and supporting details, to make inferences on the basis of the passage, to analyze the arguments used by the author, to recognize tone or attitude, or to make generalizations from information in the passage.

Questions in the mathematical sections of the test are designed to measure abilities related to college-level work in liberal arts, sciences, engineering, and other fields requiring mathematics. The tasks posed on the test are designed to assess how well students understand mathematics, how well they can apply what is already known to new situations, and how well they can use what they know to solve nonroutine problems. The test content is almost equally divided among arithmetic reasoning, algebra, and geometry; included are a few miscellaneous questions that cannot be classified in any of the three areas. For example, questions testing logical reasoning or the ability to understand and apply a new mathematical definition are classified as miscellaneous.

The mathematics questions are presented in two formats: regular multiple choice (40 questions) and quantitative comparison (20 questions). The regular multiple-choice questions are now familiar to most test-takers. The quantitative comparison questions emphasize the concepts of inequalities and estimation.

METHOD

This study is based on the secondary analysis of data gathered from nine recent administrations of the SAT from June 1986 through December 1987. This pool of information includes item statistics on 765 verbal and 540 mathematical items computed for subgroups of white, Hispanic, black, Asian American, male, and female examinees. They represent just a portion of the data produced by ETS staff through a series of generalized programs that produce a plethora of information on differential item functioning (DIF). In particular, this study focused on four reference versus focal group comparisons. Whites served as the reference group in each of three ethnic comparisons and were matched in turn with Hispanics, blacks, and Asian Americans. Inclusion in the sample depended on an affirmative response on the Student Descriptive Questionnaire (SDQ) that English was at least one of the examinees' first spoken languages. The fourth contrast featured males and females.

The first step taken to explore the relationship between DIF and item difficulty was to build a large regression model that predicts DIF from salient item characteristics. The primary independent variable was a measure of item difficulty, placed on the delta scale used extensively at ETS. The delta metric is obtained from the percentage correct through the inverse normal transformation and is scaled to have a mean of 13 and a standard deviation of 4. On this scale, delta increases for more difficult items and decreases for easier items. Since delta statistics are population-dependent, we used equating parameters to place them on a common scale across test forms. Thus, the equated delta (EQDEL) was the actual variable included in the model.

A second important item characteristic, the biserial correlation (RBIS), which estimates an item's discrimination level, was also used as an independent vari-

1. This description also appears in Schmitt and Dorans 1988, 2–4.

able in the model. The criterion used in the calculation of biserial correlations was the total scaled score for the test sections—verbal or mathematical—from which the item came.

A third variable added to the model was an index of differential omission—specifically, the standardized difference in percentage omit (DIFPOM). This measure is completely analogous to the standardized difference in percentage correct (STDP) and is discussed in Dorans, Schmitt, and Bleistein (1988). Unlike STDP, DIFPOM compares percentages of omits rather than percentages answered correctly for two matched groups. As mentioned in the Introduction, there has been some research linking differential omission to DIF (Schmitt and Bleistein 1987). In addition to the standardized difference in percentage omit, a corresponding measure of standardized difference in percentage not reaching an item (DIFPNR) was also included in the model. Note that indices comparable with these based on Mantel-Haenszel methodology were not available for this study.

We suspected that the relationship between DIF and item difficulty might depend on item type. Consequently, variables representing the four verbal (ALTYP, ANTYP, RCTYP) or two mathematical (REGMATH) item types were added to the model. Similarly, effect variables to account for differences in test form were included. Terms for estimating interactions of difficulty and either item type or test form were considered individually.

Preliminary analyses indicated that form effect variables and interaction terms added little to the model and so were dropped. The item type effect variables, on the other hand, were often significant predictors. Based on these results and previous research findings, we adopted two basic models. The first model was for use on the overall verbal or mathematical sections and the second for use on item sets of a given type.

Verbal Model

$$\text{DIF} = a_1\text{EQDEL} + a_2\text{RBIS} + a_3\text{DIFPOM} + a_4\text{DIFPNR} + a_5\text{ALTYP} + a_6\text{ANTYP} + a_7\text{RCTYP} + b$$

Mathematical Model

$$\text{DIF} = a_1\text{EQDEL} + a_2\text{RBIS} + a_3\text{DIFPOM} + a_4\text{DIFPNR} + a_5\text{REGMATH} + b$$

Item Type Level Model

$$\text{DIF} = a_1\text{EQDEL} + a_2\text{RBIS} + a_3\text{DIFPOM} + a_4\text{DIFPNR} + b$$

The dependent variable, of course, was a measure of the level of DIF in the item. Many procedures have been developed for detecting the presence of DIF. The data files for this project provided a choice between two relatively new approaches used extensively at ETS: the Mantel-Haenszel statistic (Holland and Thayer 1988) and the standardization index (Dorans and Kulick 1983, 1986), on either the delta or the percentage scale. Preliminary analyses showed correlations between indices from the two approaches exceeded .99 when placed on the same metric. These correlations were consistent across focal and reference groups for both the verbal and the mathematical sections of the test. They also exhibited similar patterns of correlations with other variables. These results confirm those found elsewhere (Wright 1987)—that these two measures of DIF are highly related. The correlation drops somewhat when the indices are on different metrics, which probably reflects that the transformation from percentage to delta is nonlinear. In fact, the Mantel-Haenszel index on the delta scale (MHD) is more highly correlated with the standardization index on the delta scale (STDD) than it is with the Mantel-Haenszel index on the percentage scale (MHP) in many cases. The standardization index on the percentage scale (STDP) is also as highly correlated with MHP as it is with STDD in most cases. Although analyses using all four independent variables were performed, only results based on using MHD are presented because choice of DIF methodology and metric appears to have only negligible effects and because ETS has operationally adopted the Mantel-Haenszel method for flagging items for differential item functioning. Results based on using STDP, STDD, and MHP as the dependent variable are available on request from the first author of this report.

The Mantel-Haenszel statistic was adapted by Holland (1985) as an approach to detecting DIF. When the total test score is used as the matching criterion, the basic data used by the Mantel-Haenszel approach are contained in the $2 \times 2 \times S$ contingency table. For each item at each score level s, data from two groups of examinees can be arranged as a 2×2 table:

	Right	Wrong	Total
Focal group	R_{fs}	W_{fs}	N_{fs}
Reference group	R_{rs}	W_{rs}	N_{rs}
Total group	R_{ts}	W_{ts}	N_{ts}

R_{fs} is the number of persons in the focal group at score level s who answered the item correctly; W_{fs} is the number in the focal group at s who answered the item incorrectly; and N_{fs}, the sum of R_{fs} and W_{fs}, is the total number in the focal group at s. R_{rs}, W_{rs}, and N_{rs} are the corresponding numbers of persons in the reference group at s. $R_{ts} = R_{rs} + R_{fs}$, $W_{ts} = W_{rs} + W_{fs}$, and $N_{ts} = N_{rs} + N_{fs}$ are the corresponding numbers of people in the total group at s.

At each score level, the Mantel-Haenszel approach uses an odds-ratio

$$\alpha_s = (R_{rs}/W_{rs}) / (R_{fs}/W_{fs}) = (R_{rs}W_{fs}) / (R_{fs}W_{rs}) \quad (1)$$

to compare the reference group with the focal group. At a given score level, s, the odds-ratio measures the advantage or disadvantage that reference group members have relative to the matched focal group members on an item. If $\alpha_s > 1$, the reference group has an advantage on the item at score level s; if $\alpha_s < 1$, the advantage lies with the focal group at score level s. The Mantel-Haenszel common-odds-ratio is a weighted average of the odds-ratios across all score levels:

$$\alpha_{MH} = \frac{\sum_{s=1}^{S} M_s \alpha_s}{\sum_{s=1}^{S} M_s} \quad (2)$$

where

$$M_s = W_{rs} R_{fs} / N_{ts} \quad (3)$$

such that

$$\alpha_{MH} = \frac{\sum_{s=1}^{S} R_{rs} W_{fs} / N_{ts}}{\sum_{s=1}^{S} R_{fs} W_{rs} / N_{ts}} \quad (4)$$

Holland (1985) shows that the α_{MH} can be converted to a difference in deltas via

$$MHD = -2.35 \ln(\alpha_{MH}) \quad (5)$$

where ln is the natural logarithm function. In this study MHD is used as the DIF statistic. Note that MHD values greater than zero indicate that relative to the reference group, the focal group performed better than expected on the item; negative values indicate the reverse.

In addition to the regression model described above, Pearson correlations and scatterplots were also examined. The results of all analyses are detailed in the results and discussion sections that follow.

RESULTS: *SAT*-VERBAL SECTIONS

Hispanics and Whites

The multiple correlation (R) for the model described in the method section on the overall SAT-verbal sections is .72. The results of this regression are summarized in Table 1. At least three conclusions are apparent from this information: first, for Hispanics and whites, item difficulty is indeed related to DIF; second, this relationship is dependent on item type; and third, differential omission is by far the most important predictor in the model. Table 2 lists the correlations between regression model variables. The positive correlation of .40 between MHD and EQDEL indicates that the focal group (Hispanics) tended to perform better than expected on more difficult items and worse than expected on less difficult items, relative to the reference group (whites). Note also that the item's biserial correlation has little association with the item's DIF level.

The DIFPNR variable is applicable to the item-type models only if items of the particular item type are located at the end of a separately timed SAT section or are close enough to the end to observe not-reached responses. This explains why DIFPNR entries for the following item-type models are, in general, empty in the tables: antonym and sentence completion items for SAT-verbal sections and quantitative comparisons for SAT-mathematical sections.

The significance of the item-type terms prompted an item-type level of analysis. The results of these regressions are also presented in Table 1. Figures 1 and 2, as well as the corresponding correlations in Table 2, show the vast difference in relationships of DIF to difficulty and differential omitting, respectively, by item type. Thus, although there seems to be a fairly strong relationship between item difficulty and DIF for the overall verbal sections, this same relationship does not hold irrespective of item type. For instance, there appears to be little or no relationship between item difficulty and DIF among reading comprehension items. Perhaps the relatively small range in item difficulty for reading comprehension items is responsible for this. (Table 1 reports that the standard deviation is lower for reading comprehension than for the other item types.) For analogies, however, the relationship appears strongest, with MHD and EQDEL correlating .55.

In fact, the model seems to work best for analogy items, where the multiple correlation is .79. In other words, 62 percent of the variance of MHD for Hispanics and whites on analogies can be accounted for by differential omission rates, equated deltas, differential not-reached rates, and item-test biserial correlations. As Table 2 shows, the correlations between MHD and DIFPOM for analogies and antonyms are −.72 and −.74, respectively. The negative correlation indicates that positive DIF levels are associated with negative differences in standardized percentage omit. Since in the calculation of DIFPOM, the conditional omit rate of the reference (white) group is subtracted from the corresponding rate in the focal (Hispanic) group, negative values of DIFPOM indicate relatively more differential omitting by the reference group. Specifically, items where Hispanics perform relatively better tend to be those items where whites omit relatively more, as hypothesized. Even for reading comprehension items, where the correlation between MHD and EQDEL is .02, the regression model yields a multiple correlation of .45 (Table 1). For each item type and for the overall

verbal sections, too, differential omit rate correlates with DIF essentially as well as, or often much better than, item difficulty does. The strength of the correlation between MHD and DIFPOM varies with item type. Note, too, that DIFPNR is also a significant predictor variable.

An alternative way to measure the practical significance or value of these independent variables is to predict the expected change in the DIF level of an item when one of the independent variables is incremented by a fixed amount. Table 1 reports that the observed mean equated delta for all 180 analogy items is 10.7, with a standard deviation of 3.1. The regression weight for EQDEL in the improved model is 0.0571. Therefore, a change in item difficulty from 1 standard deviation below the mean to 1 standard deviation above (6.2 delta units) would increase the predicted level of DIF on the Mantel-Haenszel delta difference scale by 0.35. The impact of this change is subject to personal judgment and depends on the situation. Table 3 lists the means and standard deviations of MHD for all group comparisons and all levels of analysis.

Generally speaking, to observe a change in DIF of approximately two-thirds (0.35 ÷ 0.54) of a standard deviation, as a result of changing item difficulty by 2 standard deviations, seems of only moderate impact. The predicted change on DIF among analogy items by changing the differential omit rate from 1 standard deviation below to 1 standard deviation above the mean (relatively less omitting by whites by about 4.4 percentage points) is −.64, more than a 1 standard deviation decrease in MHD (even greater differential percentage correct by whites). This change is even more pronounced for antonyms, where a similar change in the differential omit rate results in a decrease of 0.98 delta units (about 1.4 standard deviations) in the DIF measure. This predicted change in the level of differential item functioning seems quite high.

Blacks and Whites

When blacks serve as the focal group, the model yields a multiple correlation of .65 for the overall verbal sections. The significance level of the item-type factors, reported in Table 4, suggests that the differential predictive ability of the model depends on item type. Figures 3 and 4 also show how the relationship of both EQDEL and DIFPOM to MHD depends on item type. The model works best for analogies, the item type where previous research (e.g., Kulick 1984) reports the highest level of DIF. The mean MHD index for analogies (see Table 3) is −.14. This is indeed greater in magnitude than the other item types, and its sign (negative) indicates that this set of items is differentially harder for blacks than for whites. Again, as is the case for all group comparisons, the reading comprehension items exhibit a smaller range of difficulty than do the other item types.

The index of differential omission, DIFPOM, is again the single best predictor of DIF overall and for each item type; DIFPNR and RBIS are also statistically significant predictors for blacks and whites.

Table 5 displays the correlations between regression variables for both the overall and item-type level models; DIFPOM exhibits consistently higher correlations with MHD than does EQDEL.

Among analogies, the predicted change in DIF on the Mantel-Haenszel delta scale when an item's EQDEL is changed by 2 standard deviations (6.2 delta units) is 0.47—about 0.85 standard deviations. On the same set of items, an increase in DIFPOM by 2 standard deviations (4.1 percentage points) results in a predicted decrease in MHD of 0.55—about 1 standard deviation. These changes in the level of DIF appear to be fairly substantial.

Asian Americans and Whites

The model works approximately as well when the Asian American focal group is compared with a white reference group as it does for the other racial comparisons. The multiple correlation on the overall verbal sections is .68. Again, for this comparison as for the Hispanic and black focal groups, the predictive ability of the model varied by item type. (See Tables 6 and 7 and Figures 5 and 6.)

There are some features of this comparison, however, that set it apart from the others. Asian Americans omit relatively more often than do matched whites and by a larger amount than observed in the other racial comparisons. Table 6 shows the mean DIFPOM to be 1.26. As a consequence, for the first time we observe (Table 7) a positive correlation (.24) between EQDEL and DIFPOM. Table 7 also shows that the overall correlation between MHD and DIFPOM (− 37) is much weaker than in previous analyses, especially among the sentence completion items. The correlation between MHD and EQDEL (.45), however, remains comparable. Thus, DIFPOM is not the single best predictor across all models.

In this model DIFPOM can be thought of as a suppressor variable. That is, it is partially masking the relationship between MHD and EQDEL. This observation is substantiated by the partial correlation between MHD and EQDEL controlling for DIFPOM, which is .60. The Asian Americans, it seems, are doing differentially better on more difficult items despite their differential omission, not because of it.

The fit of the model is especially high for antonym item types, with a multiple correlation of .80. The model also does better among the sentence completion item types than it did for either Hispanic or black com-

parisons, despite the extraordinarily low correlation between MHD and DIFPOM (.01).

Among antonyms, the predicted change in DIF on the MHD index when an item's EQDEL is increased by 2 standard deviations (6.8 delta units) is an increase of 0.84 delta units (1.4 standard deviations). When the DIFPOM of an antonym item is increased by 2 standard deviations (4.3 percentage points), the predicted value of MHD decreases by 0.76—about 1.2 standard deviations. These changes in DIF that can be effected by changes in either EQDEL or DIFPOM seem substantial.

Females and Males

The fourth comparison placed females in the focal group and males in the reference group. The results of the regression analyses and the correlation coefficients are summarized in Tables 8 and 9 respectively. The mean DIFPOM (−.73) in Table 8 shows that females omit differentially less than do males on verbal items. The most notable result was the correlation of .01 between MHD and EQDEL (see Table 9). Despite the fact that item difficulty seems to have virtually no relationship to DIF among the females and males, the model nonetheless yielded a multiple correlation of .67, comparable with the racial group comparisons.

Although the overall verbal regression indicates item type is a significant predictor, the variance in predictive ability across item types does not vary so much as with the other group comparisons. Figure 7 reveals that the relationship between MHD and EQDEL is equally low across all item types. The scatterplots in Figure 8 indicate that the relationship between DIF and differential omission is quite strong for the female and male comparison. The item-type level model works most effectively on antonyms (multiple correlation of .74) and least well on sentence completion items (.59). Increasing the DIFPOM of an antonym item by 5.1 percentage points (2 standard deviations) decreases the predicted MHD by 1.2 delta units (about 1.5 standard deviations).

RESULTS: *SAT*-MATHEMATICAL SECTIONS

Hispanics and Whites

For the SAT-mathematical sections of the test, the model does not work nearly so well as for the SAT-verbal sections. Table 10 lists a multiple correlation of only .48. The nonsignificant item-type term indicates that the model works comparably for each of the two item types. The variables EQDEL, DIFPOM, and DIFPNR all contribute significantly to the model. As shown in Table 11, the biggest difference in correlations from the verbal analysis is that the differential omit rate correlates much more weakly with DIF (−.39 compared with −.67 in Table 2). Hispanics omit differentially more than whites do on mathematical items.

Changing either EQDEL or DIFPOM by 2 standard deviations changes the predicted MHD by less than about 0.2 delta units. Thus, the fit of the mathematical model for Hispanics and whites pales next to the verbal model.

Blacks and Whites

The drop-off in the model's predictive ability from the verbal to the mathematical sections is less dramatic for blacks than for Hispanics. Nonetheless, the multiple correlation in Table 12 is only .56 (compared with .65 in Table 4). The two item types seem to predict MHD equally well.

Blacks omit differentially more than do whites on the regular mathematical item types. The reverse is true on the quantitative comparison item types. The predicted change in DIF level resulting from change in a given independent variable is minimal in all cases. The correlations for blacks and whites (see Table 13) are similar but slightly higher than those for Hispanics and whites (see Table 11).

Asian Americans and Whites

The mathematical model works equally as well as the verbal model for this comparison, but for different reasons. Table 14 shows a multiple correlation of .68. The mathematical model relies primarily on the high correlation (−.59) between MHD and DIFPOM (see Table 15). Asian Americans continue to omit differentially more than do whites (as in the verbal sections), though not so much. The model works a little bit better on regular mathematical item types. In fact, changing the DIFPOM level of a regular mathematical item type by 2 standard deviations decreases the predicted value of MHD by 0.74 delta units.

Females and Males

The mathematical model works nearly as well as the verbal model for females and males. In Table 16 the multiple correlation is .61 (compared with .67 in Table 8). Table 17 shows that DIFPOM and MHD correlate fairly highly (−.60). The mathematical data reveal one twist, however; females omit differentially more than do males. The opposite was true for the verbal data.

A change of 2 standard deviations in DIFPOM (3.4 percentage points) results in the model's predicting a 0.60 delta unit change in MHD.

The model seems to be more effective on the regular mathematical items.

DISCUSSION

Findings from this research are consistent with the previous findings of Schmitt and Bleistein (1987), Freedle and Kostin (1987), and Freedle and Kostin (1988): item difficulty is related to DIF for some subpopulations.

This relationship appears strongest on the verbal sections, especially for Hispanic and white and black and white comparisons. The correlations are uniformly lower for the mathematical sections, with the exception of the female and male comparison, which is low in both cases. Other factors such as biserial correlations and DIFPNR contribute only minimally and inconsistently.

An even stronger relationship exists between differential omission and DIF. Correlations between DIFPOM and MHD are high in nearly all comparisons, even female and male, for both verbal and mathematical sections.

What would account for the observed strong association between differential omitting and DIF? A differential advantage on difficult items for the group that guesses differentially more might be contributing to the relationship. Consider the situation where the study groups exhibit differential omission rates; that is, one group omits an item relatively more frequently than does another group of matched ability. Since the SAT-verbal (or mathematical) score has a correction for guessing, differential omission should, on average, have no net effect on the overall SAT-verbal (or mathematical) score. That is, for a given relatively low verbal score, widely different omitting patterns are possible. Recall that the overall SAT-verbal (or mathematical) score is the matching criterion for the DIF indices. On the individual item score, however, there is no correction for guessing. Hence, a group is more likely to increase its percentage answering correctly (item score), but not its overall verbal score, if its members *guess* rather than *omit* when unsure of the answer. *The effect on item score becomes more pronounced as the proportion of the matched groups unsure of the answer (those either guessing or omitting) increases—i.e., as item difficulty increases.*

To recapitulate: as item difficulty increases, the potential for differential omission increases and, if present, in turn likely increases the observed level of DIF (as measured by the DIF indices studied). This pattern of differential omission, with Hispanics and blacks omitting less than whites do, has been reported in recent studies (Rivera and Schmitt 1988; Schmitt and Bleistein 1987) and found to be present in the current study as well. Thus, on difficult items, where more omitting and guessing are apt to take place, one might also expect greater levels of DIF favoring Hispanics and blacks relative to whites. Since DIF sums to zero (approximately) across all items in the analysis, if the more difficult items as a group display positive DIF, then the relatively easy items as a group must display negative DIF. In this way, differential omission can be offered as a partial explanation of the observed relationship between DIF and item difficulty.

Two points need to be made concerning the analyses based on differential omission. First, a slight artifact creeps into the data but seems to have little impact. By definition, the last item in a timed section cannot be omitted. If not answered, the last item is considered not reached. In fact, all unanswered items at the end of a timed section are considered not reached until one is actually responded to. Hence, two verbal items—one analogy type and one reading comprehension type—will always have differential omit rates of zero. Since these items are at the end of a section, they are presumably at least somewhat difficult—just the kind where one might expect omission. This statistical artifact should serve to suppress the correlation between DIF and DIFPOM. In parallel analyses, where the last item in each timed section was deleted, the correlation between these variables increased only slightly. In the interest of consistency, therefore, analyses were based on all items.

The second point regarding the use of differential omit rates in predicting differential functioning is the ipsative nature of the relationship between the two variables, as measured here. That is, MHD and DIFPOM are *dependent*. Observe that all response percentages must sum to 100 percent. If omitting and not reaching an item are merely considered two additional response alternatives, then the sum across all response percentages *including omits and the correct response* is 100 percent. This is true whether the response percentages are standardized (as they are here) or not. This constraint has nothing to do with DIF methodology. For a given item, if the percentage correct is greater, e.g., in the focal group than in the reference group, by a specified amount, then the sum of the response percentages for all other responses *must* be greater for the reference group than for the focal group by that same amount. Thus, if an item exhibits positive DIF (a relatively greater percentage of the focal group answered correctly), then there is necessarily some negative DIF—i.e., greater relative percentages in the reference group on at least one response alternative. There is no reason, however, to expect this negative DIF to be found in the omit rate or in any single distractor.

Consider a simpler model with just three sources of DIF. One source is the keyed response to the item. It is DIF from this source that is measured by MHD. A second source of DIF is the set of item distractors. There are four distractors on each SAT-verbal item. Third, the no-response alternative to an item is a potential source of DIF. "No response" can be further divided into omits and not reached. Since DIF from all

sources on an item must sum to zero (because of their ipsative nature), the sum from any two sources must equal the negative of the third source. But it is not true that DIF from the second and third sources individually must be opposite in sign to DIF from the first source. In fact, as Figure 6 shows, it is possible for items to display both positive DIF on the keyed response (MHD) and positive differential omission (DIFPOM). What is difficult to discern is whether it is the presence of strong differential omission that produces DIF on the keyed response, or whether it is DIF on the keyed response that manifests itself as negative DIF in the omit rates. Because of this relationship, correlations between MHD and DIFPOM may be spurious.

It should also be noted, however, that the computation of the DIF index (MHD) did not include "not reached" as a response alternative. That is, W_{fs} and W_{rs} in Equation (1) consist of those examinees in the focal and reference groups who either answered incorrectly or omitted the item. The separate calculation of DIFPOM did include not reached as a response alternative. So, for items that exhibit nonzero differential not-reached rates, the constraint of all response percentages totaling 100 percent is not strictly applicable (the relationship is not exactly ipsative). Furthermore, although the Mantel-Haenszel delta difference is closely related to the standardized difference in percentage correct, it is not identical, and thus the relationship between MHD and DIFPOM is not so strong as that for STDP and DIFPOM.

In supplemental analyses, predicting DIF from a model without DIFPOM still yielded fairly high multiple correlations, except for the female and male comparison. For example, on the SAT-verbal sections, the multiple correlations obtained from models without DIFPOM were .46, .45, .46, and .17, as opposed to .72, .65, .68, and .67 with DIFPOM, for the Hispanic and white, black and white, Asian American and white, and female and male comparisons, respectively. On the SAT-mathematical sections the multiple correlations obtained without using DIFPOM in the model were .34, .36, .31, and .17, as opposed to .48, .56, .68, and .61 with DIFPOM, for the Hispanic and white, black and white, Asian American and white, and female and male comparisons, respectively.

The role of DIFPOM seems to be more important on formula-scored tests than on rights-only-scored tests. Consider a test that is scored rights-only—i.e., there is no correction or penalty for guessing incorrectly. Clearly, on tests such as these there is no benefit in omitting, and one would expect to observe far less omitting than on a formula-scored test such as the SAT-verbal test. Further, since omitting is less frequent, differential omitting is presumably even less frequent. Therefore, on rights-only-scored tests, DIFPOM would be of little or no value in predicting DIF, and test-taking strategies (e.g., omitting) would not be likely to produce group differences. On formula-scored tests, however, different groups may be more likely to adopt different omitting or guessing strategies. Differences in these strategies might contribute to DIF on the keyed response. Certainly differential omission is going to result in counter-DIF somewhere else, whether it is on the distractors or on the key. Thus, although it is not surprising that omitting is inversely related to answering an item correctly, it seems important that observed differential omitting might be contributing to DIF and that test administration instructions are also a factor.

SUMMARY

This study examined the relationship of DIF to item difficulty as well as to other variables. The data comprise verbal and mathematical item statistics from nine recent administrations of the SAT. Based primarily on a series of correlation and regression analyses, a number of conclusions were reached. Item difficulty is related to DIF. The nature of that relationship appears to be independent of the DIF indices examined here (Mantel-Haenszel and standardization approaches). The relationship was not dependent on test form. The relationship was stronger on the verbal sections than on the mathematical sections and was substantial only for the racial comparisons, not for the female and male contrast. The relationship is such that more difficult items tend to exhibit positive DIF (DIF favors the focal group over the white reference group).

The item test biserial correlation displayed neither a consistent nor a strong relationship with DIF.

Another index, the standardized difference in percentage omit (DIFPOM), correlated very highly (negatively) with DIF. Differential omission refers to a relative difference in omit rates between groups matched in ability. In fact, DIFPOM consistently was a better predictor of DIF in most models than was item difficulty (EQDEL). The relationship between DIF and DIFPOM held up across all four comparisons, including gender. It was also present in the mathematical sections with nearly the same magnitude exhibited in the verbal sections.

Although DIF and DIFPOM are dependent measures because of their ipsative relationships, there is no reason why positive DIF must be counterbalanced with negative differential omitting, rather than negative DIF on the item's distractors. To what extent DIF is a consequence of differential omission and to what extent differential omission is a manifestation of DIF is problematic. Nonetheless, the presence of differential omission on a test has the *potential* to influence DIF indices and therefore should be an important concern. Future studies might compare rights-only-scored tests with formula-

scored tests, in terms of the level of DIF and the relationship of DIF to both differential omission and difficulty.

Among other findings is that Hispanic and black focal groups tended to omit differentially less than did the white reference groups. For Asian Americans the reverse holds. For females and males, the direction depends on the test sections. In general, groups that guess more (omit differentially less) experienced a relative advantage, as measured by the DIF indices studied here (high-positive DIF values), on the more difficult items. The Asian Americans are an exception to this finding. Asian Americans tended to omit differentially more, and yet they still experienced a relative advantage on difficult items (as seen by the correlation between MHD and EQDEL).

Differential not-reached rate exhibited a much weaker relationship to DIF than did differential omit rate.

The strength of the relationships varied across item type. Generally, in the racial comparisons the model worked best for analogy and antonym verbal item types.

Further research is needed, not only to confirm or to contest these findings, but to explore alternative explanations for the observed DIF and item difficulty relationship as well as the DIF and differential omission relationship. In particular, formula scoring of the items in the DIF analysis, consistent with the test scoring, might eliminate any effects due to differential omission. Also, DIF analyses based only on rights and wrongs, excluding omits, might provide estimates of DIF different from those obtained when omits are treated as wrong, especially for difficult items.

REFERENCES

Clemans, W. V. 1956. An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs*, No. 14.

Dorans, N. J. 1982. *Technical review of item fairness studies: 1975–1979.* ETS Research Report No. 82-90. Princeton, N.J.: Educational Testing Service.

Dorans, N. J., and E. Kulick. 1983. *Assessing unexpected differential item performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the standardization approach.* ETS Research Report No. 83-9. Princeton, N.J.: Educational Testing Service.

Dorans, N. J., and E. Kulick. 1986. Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement* 23:355–68.

Dorans, N. J., A. P. Schmitt, C. A. Bleistein. 1988. *The standardization approach to assessing differential speededness.* ETS Research Report No. 88-31. Princeton, N.J.: Educational Testing Service.

Freedle, R., and I. Kostin. 1987. *Semantic and structural factors affecting the performance of matched black and white examinees on analogies items from the Scholastic Aptitude Test.* Princeton, N.J.: Educational Testing Service. Research Report. Final Report, PRPC project, submitted August 1987.

Freedle, R., and I. Kostin. 1988. *Relationship between item characteristics and an index of differential item functioning (DIF) for the four GRE verbal item types.* ETS Research Report No. 88-29. Princeton, N.J.: Educational Testing Service.

Holland, P. W. 1985. On the study of differential item performance without IRT. Paper presented at annual meeting of the Military Testing Association, San Diego, Calif.

Holland, P. W., and D. T. Thayer. 1988. Differential item functioning and the Mantel-Haenszel procedure. In *Test Validity*, ed. H. Wainer and H. I. Braun. Hillsdale, N.J.: Erlbaum.

Kulick, E. 1984. *Assessing unexpected differential item performance of black candidates on SAT form CSA6 and TSWE form E33.* ETS Statistical Report No. 84-80. Princeton, N.J.: Educational Testing Service.

Rivera, C., and A. Schmitt. 1988. *A comparison of Hispanic and white students' omit patterns on the Scholastic Aptitude Test.* ETS Research Report No. 88-44. Princeton, N.J.: Educational Testing Service.

Rogers, H. J., and E. Kulick. 1987. An investigation of unexpected differences in item performance between blacks and whites taking the SAT. In *Differential item functioning on the Scholastic Aptitude Test*, ed. A. P. Schmitt and N. J. Dorans. ETS Research Memorandum No. 87-1. Princeton, N.J.: Educational Testing Service.

Rogers, H. J., N. J. Dorans, and A. P. Schmitt. 1986. *Assessing unexpected differential item performance of black candidates on SAT form 3GSA08 and TSWE form E43.* ETS Statistical Report No. 86-22. Princeton, N.J.: Educational Testing Service.

Schmitt, A. P., and C. A. Bleistein. 1987. *Factors affecting differential item functioning for black examinees on Scholastic Aptitude Test analogy items.* ETS Research Report No. 87-23. Princeton, N.J.: Educational Testing Service.

Schmitt, A. P., and N. J. Dorans. 1988. *Differential item functioning for minority examinees on the SAT.* ETS Research Report No. 88-32. Princeton, N.J.: Educational Testing Service.

Taking the SAT: A guide to the Scholastic Aptitude Test. 1988. New York: College Entrance Examination Board.

Wright, D. 1987. An empirical comparison of the Mantel-Haenszel and standardization methods of detecting differential item performance. In *Differential item functioning on the Scholastic Aptitude Test*, ed. A. P. Schmitt and N. J. Dorans. ETS Research Memorandum No. 87-1. Princeton, N.J.: Educational Testing Service.

FIGURES

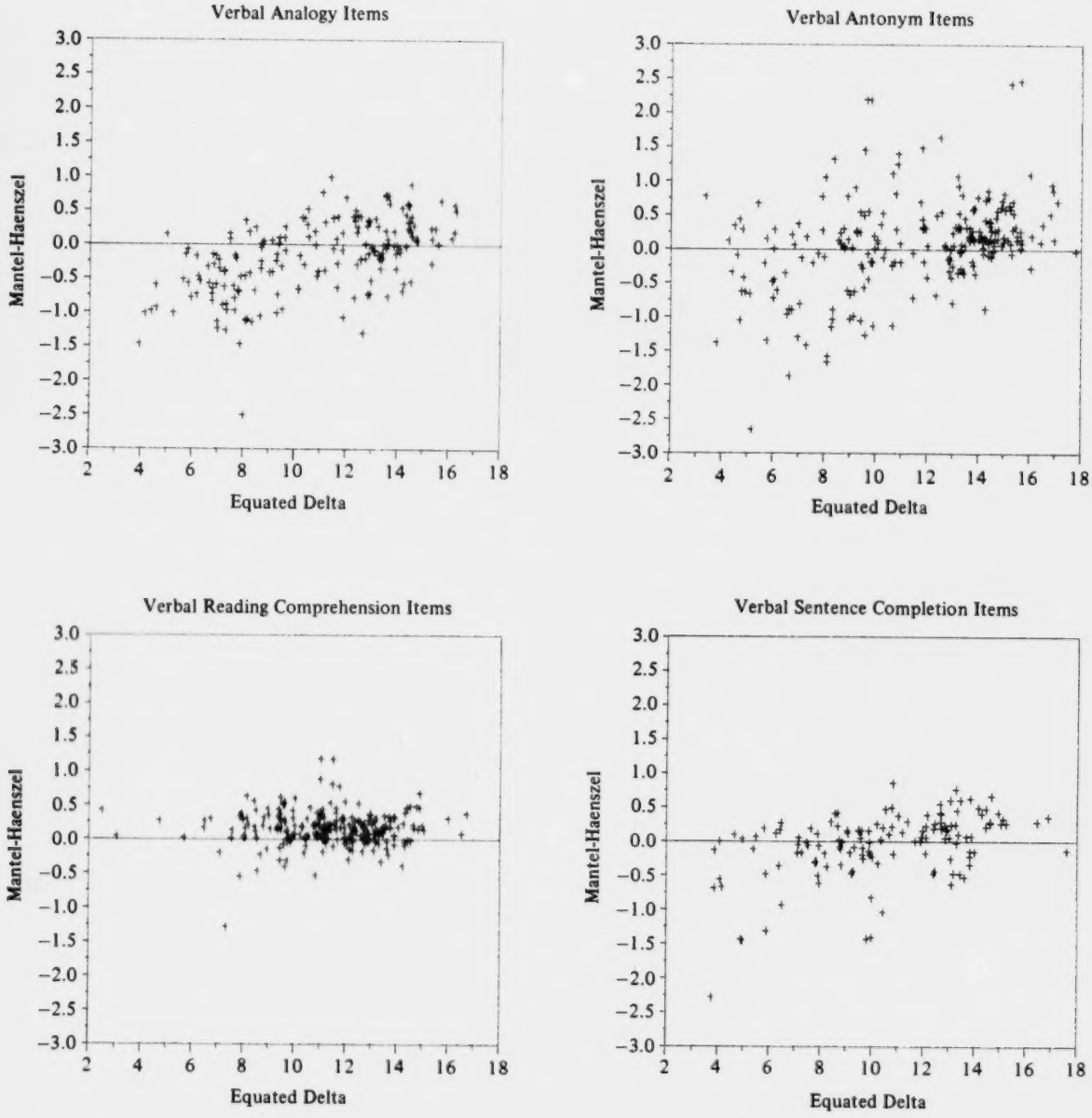

Figure 1. Mantel-Haenszel Delta-Difference versus Equated Delta (Hispanics and Whites)

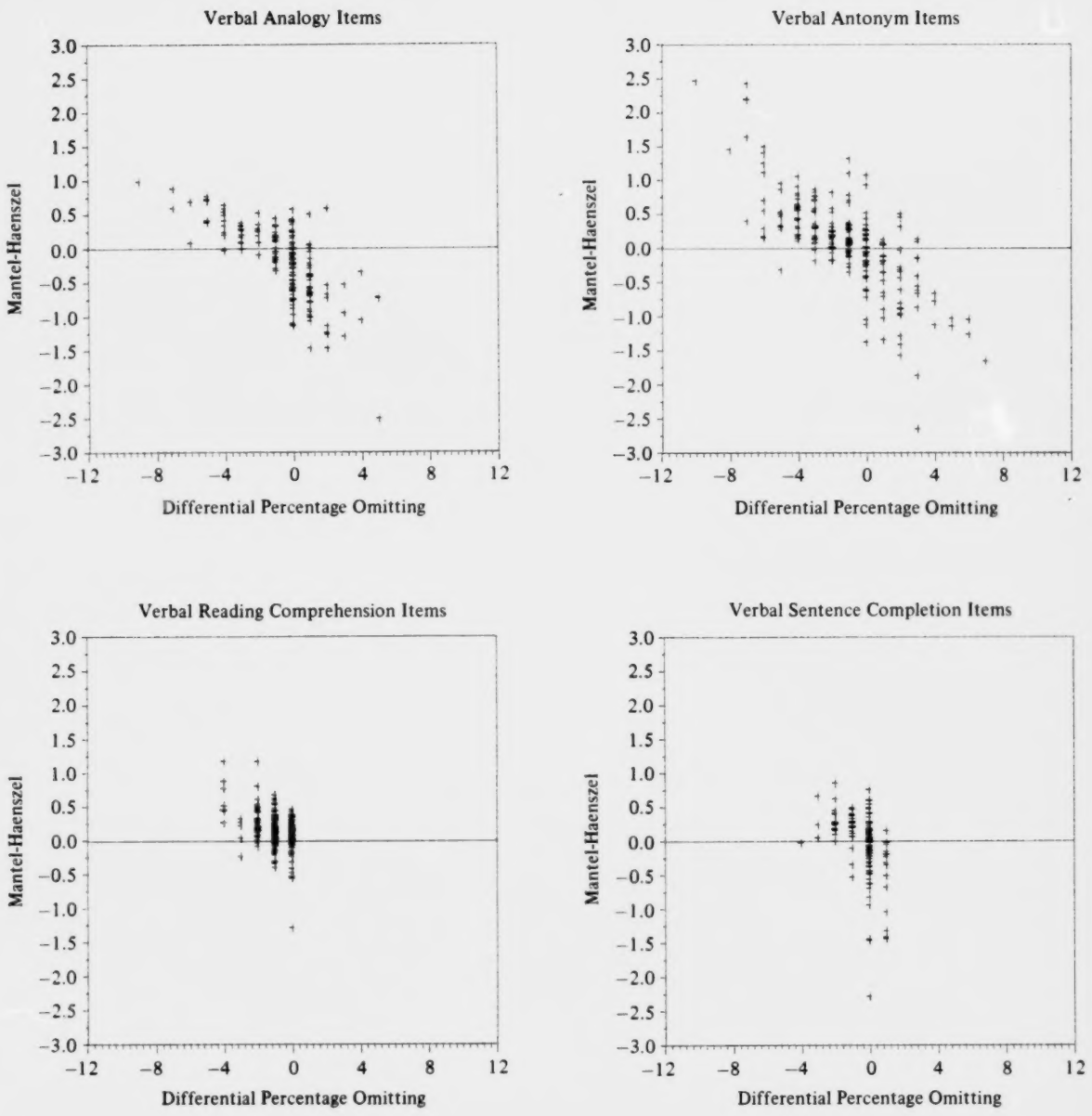

Figure 2. Mantel-Haenszel Delta-Difference versus Differential Percentage Omitting (Hispanics and Whites)

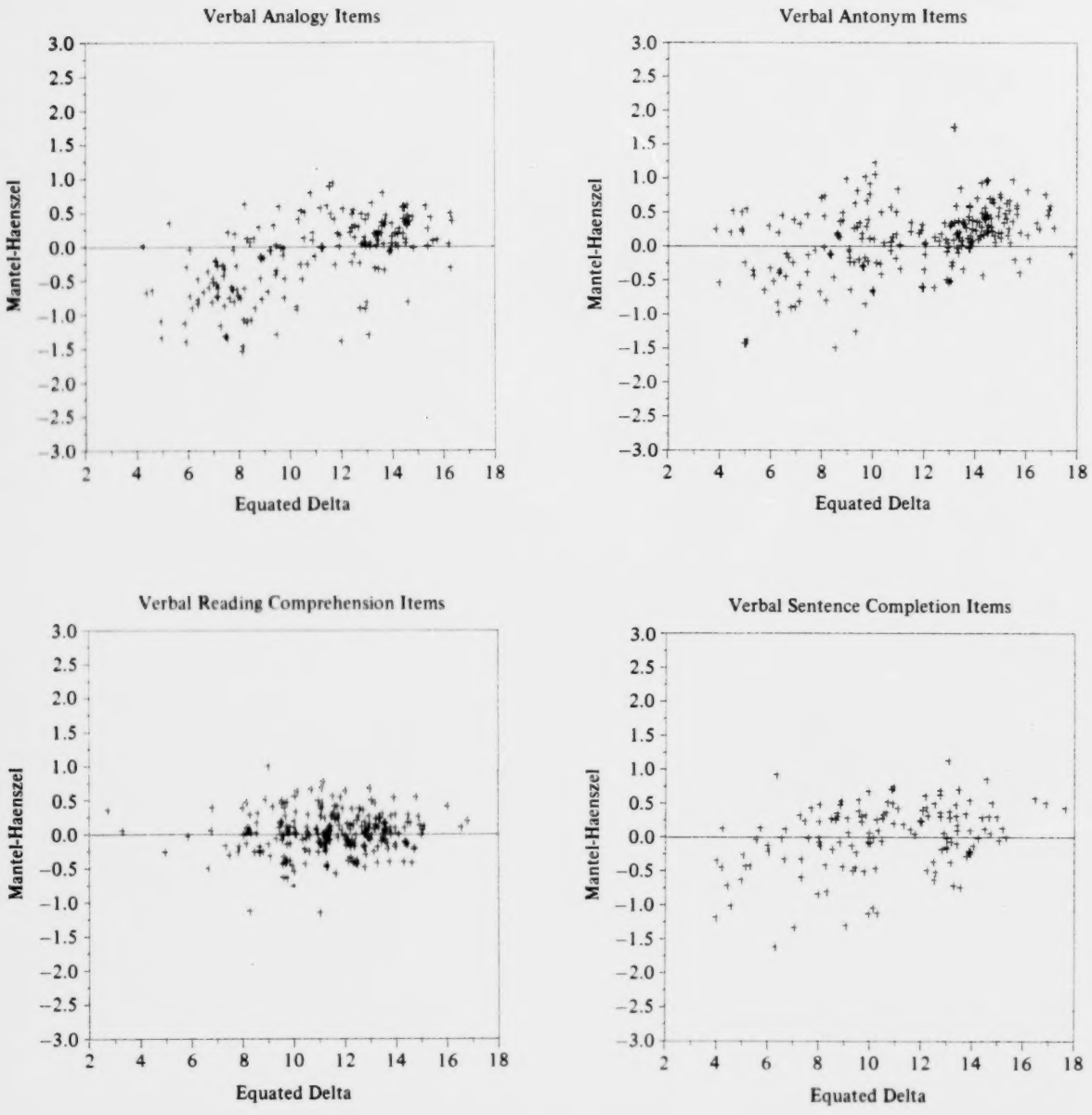

Figure 3. Mantel-Haenszel Delta-Difference versus Equated Delta (Blacks and Whites)

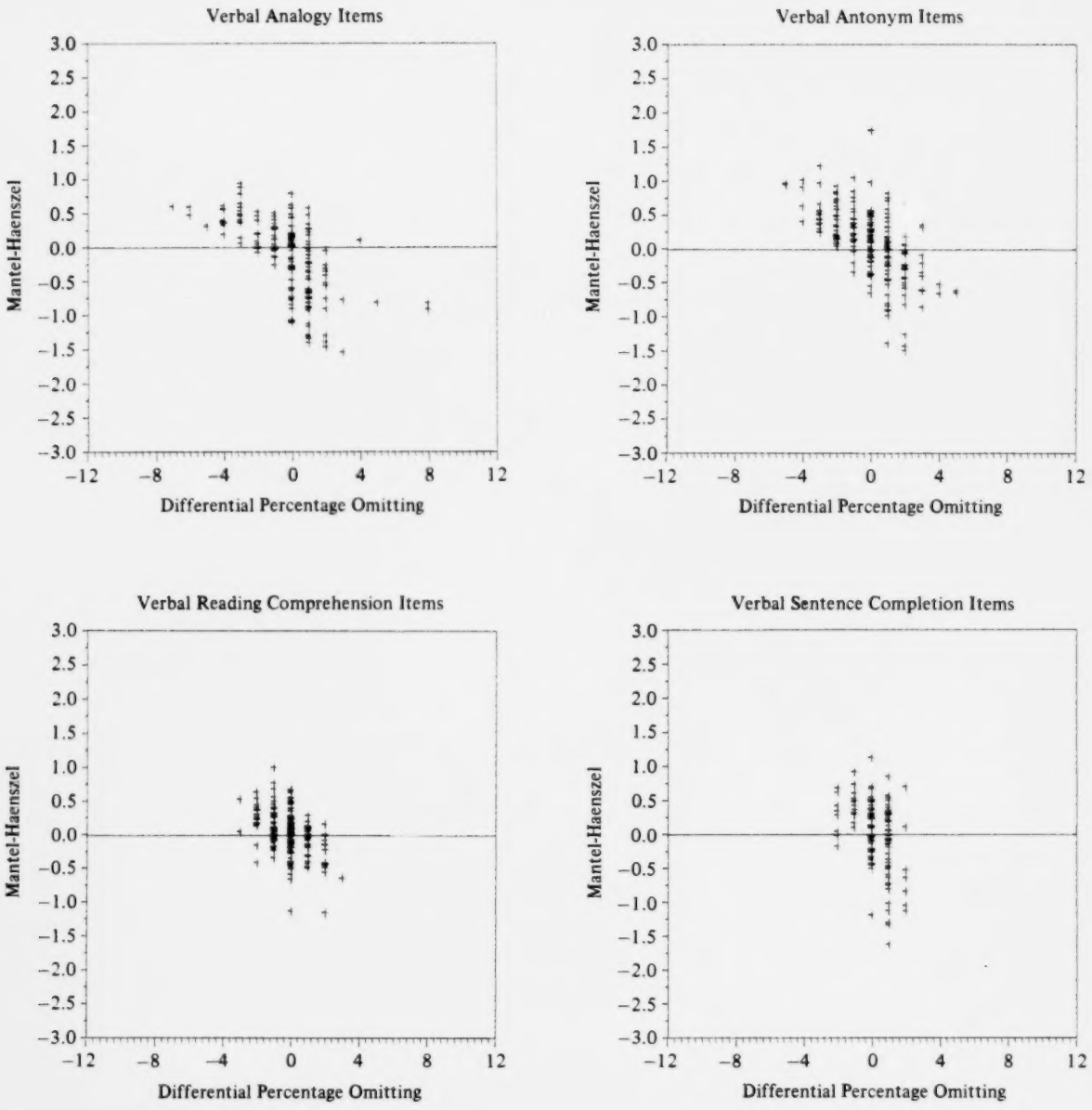

Figure 4. Mantel-Haenszel Delta-Difference versus Differential Percentage Omitting (Blacks and Whites)

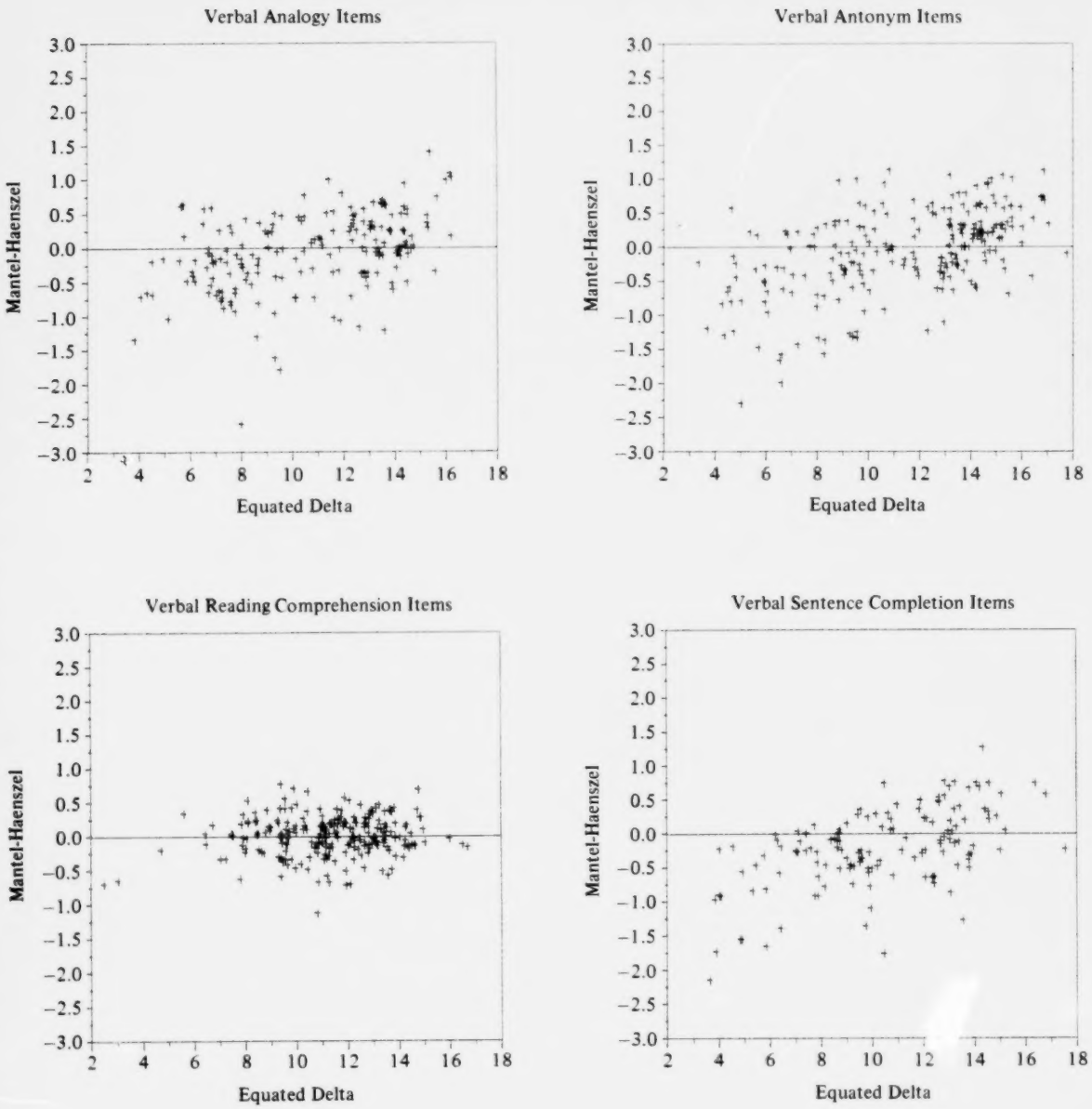

Figure 5. Mantel-Haenszel Delta-Difference versus Equated Delta (Asian Americans and Whites)

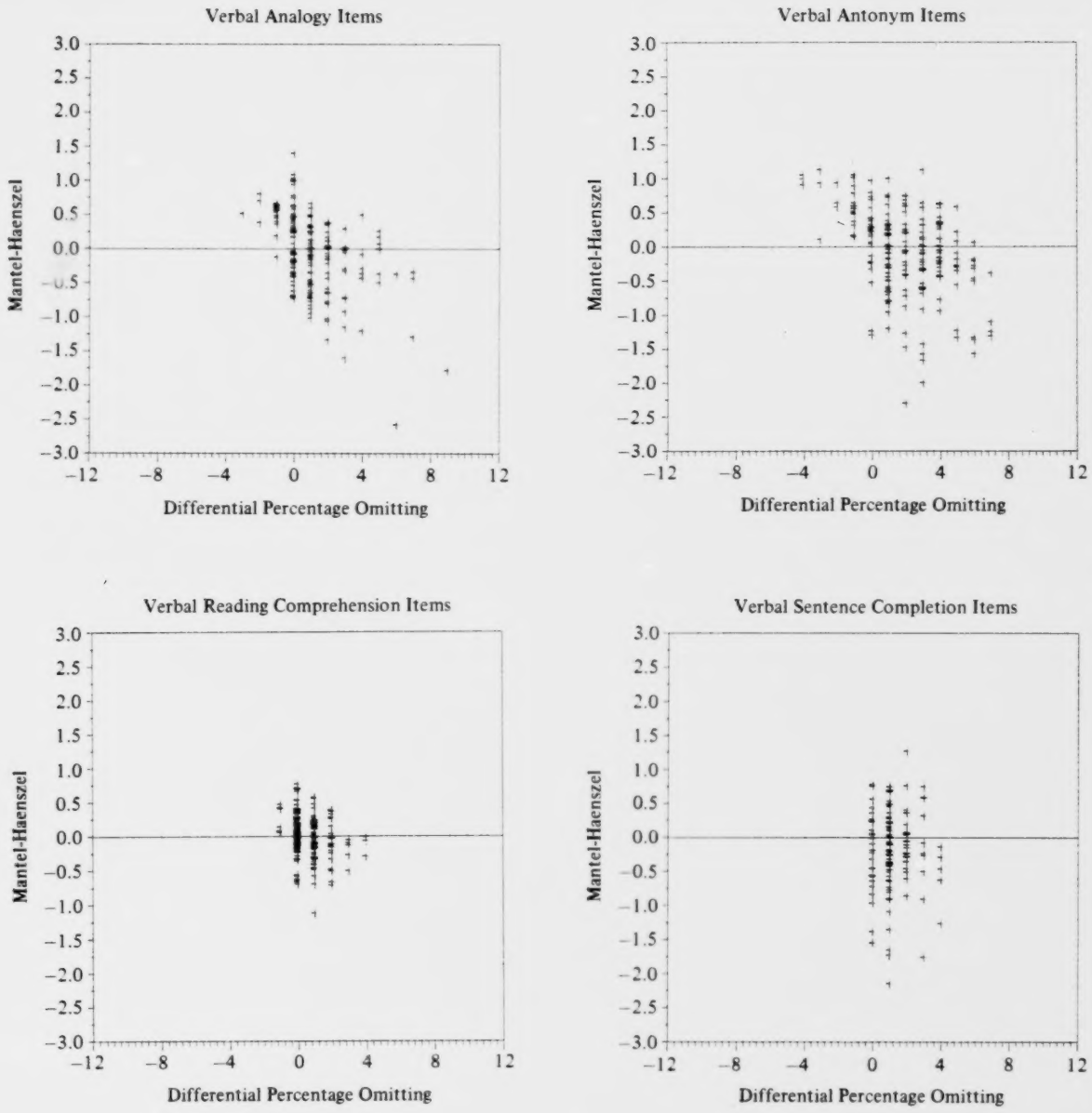

Figure 6. Mantel-Haenszel Delta-Difference versus Differential Percentage Omitting (Asian Americans and Whites)

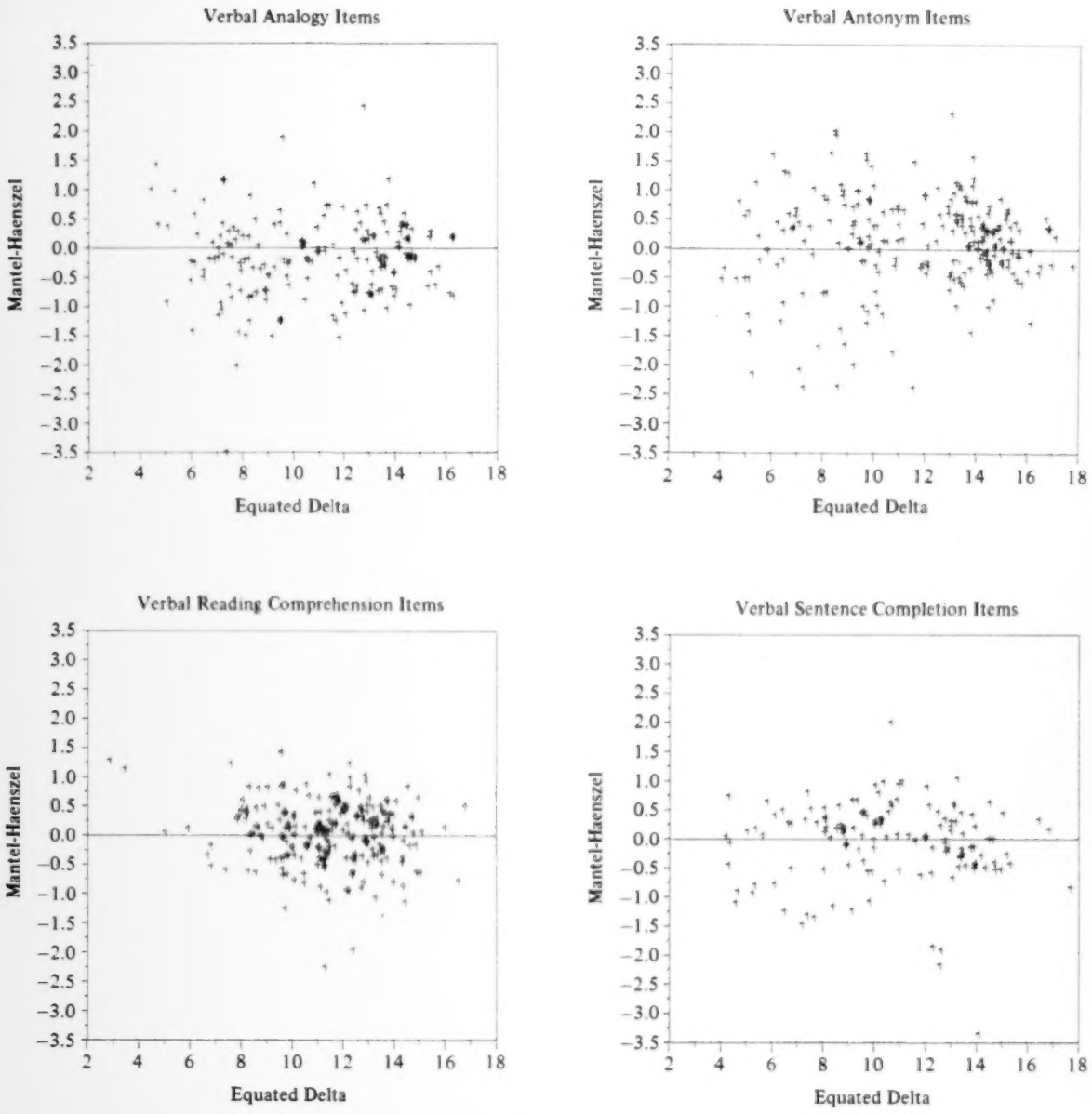

Figure 7. Mantel-Haenszel Delta-Difference versus Equated Delta (Females and Males)

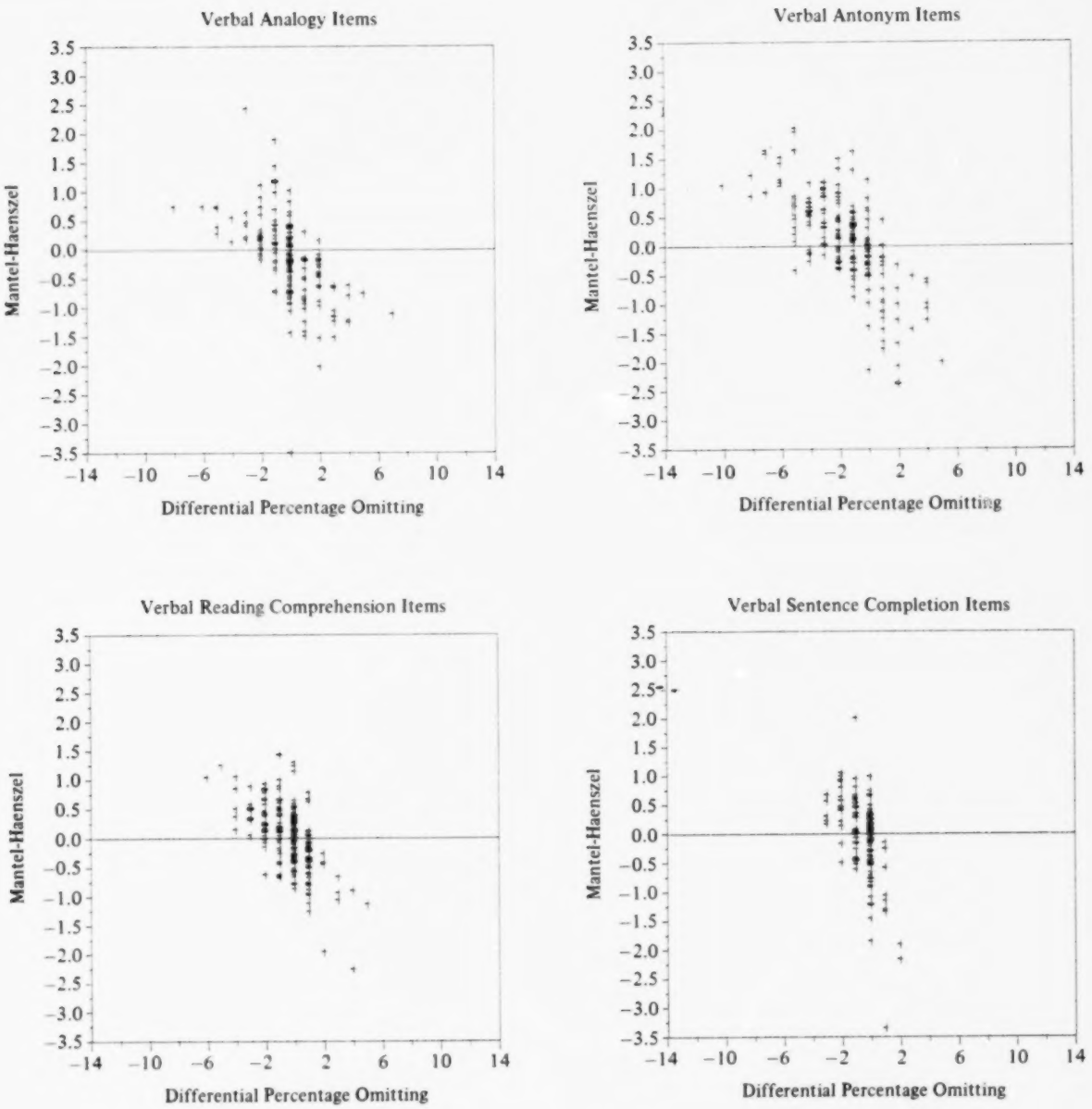

Figure 8. Mantel-Haenszel Delta-Difference versus Differential Percentage Omitting (Females and Males)

TABLES

Table 1. Regression Results for Predicting MHD from SAT-Verbal Item Characteristics (Hispanics and Whites)

Overall Verbal Sections	*Multiple R = .72*		*R Squared = .52*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 757)*	*Mean*	*SD*
EQDEL	0.0287	.005	5.86	11.0	3.0
RBIS	−0.0333	.142	− 0.23	0.47	0.10
DIFPOM	−0.1614	.007	−21.89	− 0.83	1.99
DIFPNR	−0.0308	.011	− 2.92	0.65	1.42
ALTYP	−0.1319	.025	− 5.21	0.06	0.64
ANTYP	−0.0192	.024	− 0.82	0.12	0.68
RCTYP	0.1330	.023	5.68	0.12	0.68
Analogy Item Type	*Multiple R = .79*		*R Squared = .62*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 175)*	*Mean*	*SD*
EQDEL	0.0571	.009	6.52	10.7	3.1
RBIS	0,0877	.260	0.34	0.45	0.10
DIFPOM	−0.1443	.012	−11.84	− 0.56	2.22
DIFPNR	−0.0281	.013	− 2.15	1.26	1.95
Antonym Item Type	*Multiple R = .75*		*R Squared = .56*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 221)*	*Mean*	*SD*
EQDEL	0.0213	.010	2.12	11.4	3.4
RBIS	0.2659	.310	0.86	0.47	0.11
DIFPOM	−0.1758	.012	−14.55	− 1.20	2.78
DIFPNR	—	—	—	0.0	0.0
Reading Comprehension Item Type	*Multiple R = .45*		*R Squared = .20*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 220)*	*Mean*	*SD*
EQDEL	−0.0156	.008	− 2.05	11.3	2.3
RBIS	−0.5054	.201	− 2.51	0.46	0.08
DIFPOM	−0.1355	.018	− 7.39	− 0.96	0.93
DIFPNR	−0.0200	.010	− 1.96	1.20	1.60
Sentence Completion Item Type	*Multiple R = .51*		*R Squared = .26*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 131)*	*Mean*	*SD*
EQDEL	0.0502	.013	3.97	10.4	3.2
RBIS	−0.2247	.357	− 0.63	0.52	0.10
DIFPOM	−0.1342	.041	− 3.24	− 0.34	0.96
DIFPNR	—	—	—	0.0	0.0

Table 2. Correlations Based on SAT-Verbal Data for Hispanic Focal Group and White Reference Group

Overall Verbal Sections

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.40	1.00			
RBIS	− .09	− .21	1.00		
DIFPOM	− .67	− .37	.10	1.00	
DIFPNR	− .07	− .01	− .07	.00	1.00
ALTYP	− .11	.02	− .21	− .02	.28
ANTYP	.09	.10	− .13	− .15	− .08
RCTYP	.15	.09	− .19	− .09	.29

Analogy Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.55	1.00			
RBIS	− .01	− .17	1.00		
DIFPOM	− .72	− .35	− .08	1.00	
DIFPNR	− .17	− .13	.15	.04	1.00

Antonym Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.38	1.00			
RBIS	− .18	− .35	1.00		
DIFPOM	− .74	− .40	.26	1.00	
DIFPNR	—	—	—	—	—

Reading Comprehension Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.02	1.00			
RBIS	− .05	− .21	1.00		
DIFPOM	− .40	− .29	− .13	1.00	
DIFPNR	− .03	.15	− .14	− .19	1.00

Sentence Completion Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.45	1.00			
RBIS	− .02	.07	1.00		
DIFPOM	− .42	− .44	− .00	1.00	
DIFPNR	—	—	—	—	—

Table 3. Means and Standard Deviations of Mantel-Haenszel Delta Difference Index (MHD)

Focal and Reference Group	*Verbal Mean (SD)*		*Mathematical Mean (SD)*	
Hispanic and white	Verbal	.0182 (.5339)	Mathematical	−.0215 (.2948)
	Analogies	−.1899 (.5403)	Reg. math.	.0264 (.3119)
	Antonyms	.0841 (.6857)	Quant. comp.	.0116 (.2567)
	Reading comp.	.1582 (.2622)		
	Sentence comp.	−.0475 (.4755)		
Black and white	Verbal	.0128 (.4764)	Mathematical	.0149 (.4579)
	Analogies	−.1437 (.5525)	Reg. math.	−.0012 (.4748)
	Antonyms	.1260 (.5050)	Quant. comp.	.0472 (.4203)
	Reading comp.	.0301 (.3075)		
	Sentence comp.	.0039 (.4914)		
Asian American and white	Verbal	−.0673 (.5279)	Mathematical	.0034 (.5507)
	Analogies	−.0625 (.5703)	Reg. math.	.0171 (.5942)
	Antonyms	−.0609 (.6174)	Quant. comp.	−.0239 (.4503)
	Reading comp.	.0049 (.2919)		
	Sentence comp.	.2044 (.5835)		
Female and male	Verbal	.0041 (.6978)	Mathematical	−.0120 (.4919)
	Analogies	−.1584 (.7017)	Reg. math.	−.0006 (.5180)
	Antonyms	.1158 (.7951)	Quant. comp.	−.0349 (.4340)
	Reading comp.	.0509 (.5499)		
	Sentence comp.	−.0433 (.6952)		

Table 4. Regression Results for Predicting MHD from SAT-Verbal Item Characteristics (Blacks and Whites)

Overall Verbal Sections	*Multiple R = .65*		*R Squared = .42*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 757)*	*Mean*	*SD*
EQDEL	0.0443	.005	9.35	11.1	2.97
RBIS	−0.3929	.137	− 2.86	0.48	0.10
DIFPOM	−0.1452	.009	−16.73	− 0.08	1.55
DIFPNR	−0.0377	.008	− 4.85	1.11	2.00
ALTYP	−0.1397	.025	− 5.63	0.06	0.64
ANTYP	0.0614	.023	2.62	0.12	0.68
RCTYP	0.0280	.024	1.18	0.12	0.68
Analogy Item Type	*Multiple R = .75*		*R Squared = .56*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 175)*	*Mean*	*SD*
EQDEL	0.0754	.010	7.72	10.8	3.1
RBIS	−0.0855	.283	− 0.30	0.46	0.10
DIFPOM	−0.1337	.014	− 9.53	− 0.26	2.06
DIFPNR	−0.0332	.013	− 2.65	1.96	2.20
Antonym Item Type	*Multiple R = .62*		*R Squared = .39*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 220)*	*Mean*	*SD*
EQDEL	0.0426	.009	4.80	11.5	3.4
RBIS	−0.1493	.266	− 0.56	0.48	0.11
DIFPOM	−0.1349	.015	− 8.94	0.05	1.81
DIFPNR	—	—	—	0.0	0.0
Reading Comprehension Item Type	*Multiple R = .56*		*R Squared = .32*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 131)*	*Mean*	*SD*
EQDEL	0.0006	.008	0.08	11.4	2.2
RBIS	−0.8089	.213	− 3.79	0.47	0.08
DIFPOM	−0.1629	.018	− 9.26	− 0.19	1.02
DIFPNR	−0.0354	.008	− 4.74	2.22	2.44
Sentence Completion Item Type	*Multiple R = .56*		*R Squared = .31*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 221)*	*Mean*	*SD*
EQDEL	0.0542	.012	4.70	10.6	3.1
RBIS	−0.8647	.351	− 2.46	0.53	0.10
DIFPOM	−0.1914	.039	− 4.90	0.28	0.92
DIFPNR	—	—	—	0.0	0.0

Table 5. Correlations Based on SAT-Verbal Data for Black Focal Group and White Reference Group

Overall Verbal Sections

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	—				
EQDEL	.40	—			
RBIS	− .18	− .29	—		
DIFPOM	− .51	− .20	.11	—	
DIFPNR	− .13	.06	− .10	− .12	—
ALTYP	− .12	.02	− .19	− .11	.31
ANTYP	.11	.10	− .14	− .05	−.10
RCTYP	.02	.09	− .19	− .09	.39

Analogy Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.57	1.00			
RBIS	− .14	− .27	1.00		
DIFPOM	− .62	− .29	.01	1.00	
DIFPNR	− .13	− .02	.05	− .02	1.00

Antonym Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.40	1.00			
RBIS	− .23	− .43	1.00		
DIFPOM	− .55	− .21	.16	1.00	
DIFPNR	—	—	—	—	—

Reading Comprehension Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.06	1.00			
RBIS	− .19	− .27	1.00		
DIFPOM	− .46	− .11	.00	1.00	
DIFPNR	− .10	.22	− .11	− .29	1.00

Sentence Completion Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.36	1.00			
RBIS	− .23	.01	1.00		
DIFPOM	− .41	− .06	.15	1.00	
DIFPNR	—	—	—	—	—

Table 6. Regression Results for Predicting MHD from SAT-Verbal Item Characteristics (Asian Americans and Whites)

Overall Verbal Sections	*Multiple R = .68*		*R Squared = .47*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 757)*	*Mean*	*SD*
EQDEL	0.1007	.005	20.67	11.0	3.0
RBIS	0.1194	.148	0.81	0.48	0.10
DIFPOM	−0.1690	.009	−18.88	1.26	1.70
DIFPNR	−0.0590	.025	− 2.35	− 0.00	0.57
ALTYP	0.0253	.026	0.99	0.06	0.64
ANTYP	0.0962	.024	4.01	0.12	0.68
RCTYP	−0.0495	.024	− 2.02	0.12	0.68
Analogy Item Type	*Multiple R = .71*		*R Squared = .51*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 175)*	*Mean*	*SD*
EQDEL	0.0913	.010	9.34	10.7	3.2
RBIS	0.5350	.308	1.74	0.45	0.10
DIFPOM	−0.1738	.016	−10.68	1.19	1.87
DIFPNR	−0.0664	.034	− 1.95	− 0.18	0.88
Antonym Item Type	*Multiple R = .80*		*R Squared = .64*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 220)*	*Mean*	*SD*
EQDEL	0.1238	.008	15.57	11.3	3.4
RBIS	0.0778	.254	0.31	0.48	0.10
DIFPOM	−0.1777	.012	−14.86	1.96	2.16
DIFPNR	—	—	—	0.0	0.0
Reading Comprehension Item Type	*Multiple R = .31*		*R Squared = .10*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 131)*	*Mean*	*SD*
EQDEL	0.0226	.009	2.53	11.3	2.3
RBIS	−0.1881	.232	− 0.81	0.46	0.08
DIFPOM	−0.1015	.022	− 4.63	0.61	0.91
DIFPNR	−0.0116	.029	− 0.40	0.13	0.66
Sentence Completion Item Type	*Multiple R = .64*		*R Squared = .41*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 221)*	*Mean*	*SD*
EQDEL	0.1363	.015	9.37	10.4	3.1
RBIS	−0.1964	.410	− 0.48	0.52	0.10
DIFPOM	−0.2085	.047	− 4.45	1.27	0.99
DIFPNR	—	—	—	0.0	0.0

Table 7. Correlations Based on SAT-Verbal Data for Asian American Focal Group and White Reference Group

Overall Verbal Sections

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	—				
EQDEL	.45	—			
RBIS	− .07	− .19	—		
DIFPOM	− .37	.24	− .06	—	
DIFPNR	− .03	.03	− .01	− .05	—
ALTYP	.08	.02	− .21	− .02	− .12
ANTYP	.07	.10	− .12	.18	.00
RCTYP	.13	.09	− .19	− .17	.10

Analogy Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.41	1.00			
RBIS	.04	− .15	1.00		
DIFPOM	− .50	.14	− .04	1.00	
DIFPNR	− .12	− .03	.01	− .00	1.00

Antonym Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.52	1.00			
RBIS	− .15	− .34	1.00		
DIFPOM	− .45	.26	− .11	1.00	
DIFPNR	—	—	—	—	—

Reading Comprehension Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.09	1.00			
RBIS	− .06	− .21	1.00		
DIFPOM	− .26	.30	− .08	1.00	
DIFPNR	.04	.11	− .09	− .14	1.00

Sentence Completion Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.56	1.00			
RBIS	.09	.08	1.00		
DIFPOM	.01	.49	− .18	1.00	
DIFPNR	—	—	—	—	—

Table 8. Regression Results for Predicting MHD from SAT-Verbal Item Characteristics (Females and Males)

Overall Verbal Sections	*Multiple R = .67*		*R Squared = .44*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 757)*	*Mean*	*SD*
EQDEL	−0.0415	.007	− 5.95	11.2	2.94
RBIS	−0.6225	.198	− 3.14	0.48	0.10
DIFPOM	−0.2372	.010	−23.33	− 0.73	2.01
DIFPNR	−0.1086	.025	− 4.37	− 0.21	0.81
ALTYP	−0.0226	.035	− 0.66	0.06	0.64
ANTYP	−0.0863	.033	− 2.59	0.12	0.68
RCTYP	0.0813	.034	2.42	0.12	0.68
Analogy Item Type	*Multiple R = .60*		*R Squared = .36*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 175)*	*Mean*	*SD*
EQDEL	−0.0159	.015	− 1.07	10.9	3.0
RBIS	−0.8801	.442	− 1.99	0.47	0.10
DIFPOM	−0.2005	.022	− 9.24	− 0.08	1.97
DIFPNR	−0.1041	.049	− 2.14	− 0.11	0.89
Antonym Item Type	*Multiple R = .74*		*R Squared = .55*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 221)*	*Mean*	*SD*
EQDEL	−0.0502	.013	− 3.95	11.5	3.3
RBIS	−0.4876	.361	− 1.35	0.48	0.11
DIFPOM	−0.2433	.015	−16.32	− 1.71	2.53
DIFPNR	—	—	—	0.0	0.0
Reading Comprehension Item Type	*Multiple R = .70*		*R Squared = .49*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 220)*	*Mean*	*SD*
EQDEL	−0.0738	.013	− 5.72	11.4	2.2
RBIS	−0.8219	.331	− 2.49	0.47	0.08
DIFPOM	−0.2350	.018	−13.10	− 0.39	1.50
DIFPNR	−0.1169	.023	− 5.02	− 0.63	1.16
Sentence Completion Item Type	*Multiple R = .59*		*R Squared = .35*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 131)*	*Mean*	*SD*
EQDEL	−0.0532	.017	− 3.13	10.6	3.1
RBIS	−0.1632	.489	− 0.33	0.53	0.10
DIFPOM	−0.4512	.055	− 8.19	− 0.53	0.96
DIFPNR	—	—	—	0.0	0.0

Table 9. Correlations Based on SAT-Verbal Data for Female Focal Group and Male Reference Group

Overall Verbal Sections

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	0.01	1.00			
RBIS	−0.08	−0.31	1.00		
DIFPOM	−0.62	− .22	.04	1.00	
DIFPNR	−0.13	− .11	.09	− .02	1.00
ALTYP	−0.07	.02	− .19	− .09	− .02
ANTYP	0.08	.10	− .14	− .24	.05
RCTYP	0.05	.09	− .20	.05	− .30

Analogy Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.04	1.00			
RBIS	− .14	− .30	1.00		
DIFPOM	− .57	− .09	.01	1.00	
DIFPNR	− .19	− .14	.17	.08	1.00

Antonym Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	.05	1.00			
RBIS	− .03	− .45	1.00		
DIFPOM	− .72	− .29	.08	1.00	
DIFPNR	—	—	—	—	—

Reading Comprehension Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	− .13	1.00			
RBIS	− .06	− .28	1.00		
DIFPOM	− .61	− .13	.02	1.00	
DIFPNR	− .21	− .18	.02	.02	1.00

Sentence Completion Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	− .03	1.00			
RBIS	− .13	− .02	1.00		
DIFPOM	− .55	− .33	.18	1.00	
DIFPNR	—	—	—	—	—

Table 10. Regression Results for Predicting MHD from SAT-Mathematical Item Characteristics (Hispanics and Whites)

Overall Mathematical Sections	*Multiple R = .48*		*R Squared = .23*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 534)*	*Mean*	*SD*
EQDEL	0.0313	.005	6.95	12.0	3.2
RBIS	0.1627	.104	1.57	0.55	0.11
DIFPOM	−0.0814	.009	−8.69	− 0.61	1.31
DIFPNR	−0.0334	.009	−3.85	0.93	1.77
REGMATH	0.0002	.013	0.02	0.33	0.94
Regular Mathematical Item Type	*Multiple R = .48*		*R Squared = .23*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 355)*	*Mean*	*SD*
EQDEL	0.0300	.006	5.04	12.1	3.4
RBIS	0.2552	.134	1.90	0.56	0.11
DIFPOM	−0.0824	.011	−7.58	− 0.77	1.46
DIFPNR	−0.0311	.010	−3.12	1.32	2.05
Quantitative Comparison Item Type	*Multiple R = .47*		*R Squared = .22*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 175)*	*Mean*	*SD*
EQDEL	0.0328	.007	4.59	11.8	2.9
RBIS	−0.0235	.164	−0.14	0.53	0.11
DIFPOM	−0.0823	.020	−4.03	− 0.30	0.87
DIFPNR	−0.0485	.052	−0.93	0.16	0.40

Table 11. Correlations Based on SAT-Mathematical Data for Hispanic Focal Group and White Reference Group

			Overall Mathematical Sections			
	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	*REGMATH*
MHD	1.00					
EQDEL	.33	1.00				
RBIS	.01	− .19	1.00			
DIFPOM	− .39	− .33	.06	1.00		
DIFPNR	.13	.60	− .16	− .38	1.00	
REGMATH	.01	.04	.15	− .17	.31	1.00
			Regular Mathematical Item Type			
	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	
MHD	1.00					
EQDEL	.31	1.00				
RBIS	.03	− .19	1.00			
DIFPOM	− .41	− .36	.14	1.00		
DIFPNR	.14	.68	− .24	− .38	1.00	
			Quantitative Comparison Item Type			
	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	
MHD	1.00					
EQDEL	.39	1.00				
RBIS	− .05	− .23	1.00			
DIFPOM	− .34	− .21	− .07	1.00		
DIFPNR	.17	.53	− .30	− .18	1.00	

Table 12. Regression Results for Predicting MHD from SAT-Mathematical Item Characteristics (Blacks and Whites)

Overall Mathematical Sections	*Multiple R = .56*		*R Squared = .31*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 534)*	*Mean*	*SD*
EQDEL	0.0518	.007	7.53	12.1	3.1
RBIS	−0.0349	.151	− 0.23	0.55	0.12
DIFPOM	−0.1162	.010	−11.98	− 0.14	1.81
DIFPNR	−0.0323	.010	− 2.98	1.29	2.37
REGMATH	−0.0579	.019	− 3.33	0.55	0.12
Regular Mathematical Item Type	*Multiple R = .56*		*R Squared = .32*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 355)*	*Mean*	*SD*
EQDEL	0.0456	.009	5.01	12.2	3.3
RBIS	0.0510	.192	0.27	0.57	0.11
DIFPOM	−0.1150	.016	− 9.89	− 0.44	1.90
DIFPNR	−0.0264	.011	− 2.41	1.83	2.73
Quantitative Comparison Item Type	*Multiple R = .55*		*R Squared = .31*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 175)*	*Mean*	*SD*
EQDEL	0.0604	.012	5.03	11.9	2.8
RBIS	−0.1921	.255	− 0.75	0.53	0.11
DIFPOM	−0.1319	.019	− 6.80	0.47	1.42
DIFPNR	−0.0232	.077	− 0.30	0.21	0.44

Table 13. Correlations Based on SAT-Mathematical Data for Black Focal Group and White Reference Group

			Overall Mathematical Sections			
	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	*REGMATH*
MHD	1.00					
EQDEL	.35	1.00				
RBIS	− .09	− .27	1.00			
DIFPOM	− .46	− .22	.00	1.00		
DIFPNR	.16	.62	− .21	− .31	1.00	
REGMATH	− .05	.05	.15	− .24	.32	1.00
			Regular Mathematical Item Type			
	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	
MHD	1.00					
EQDEL	.36	1.00				
RBIS	− .09	− .27	1.00			
DIFPOM	− .52	− .33	.14	1.00		
DIFPNR	.20	.71	− .31	− .29	1.00	
			Quantitative Comparison Item Type			
	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	
MHD	1.00					
EQDEL	.34	1.00				
RBIS	− .06	− .30	1.00			
DIFPOM	− .38	.13	− .23	1.00		
DIFPNR	.23	.60	− .33	− .00	1.00	

Table 14. Regression Results for Predicting MHD from SAT-Mathematical Item Characteristics (Asian Americans and Whites)

Overall Mathematical Sections	*Multiple R = .68*		*R Squared = .46*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 534)*	*Mean*	*SD*
EQDEL	0.0563	.006	9.56	11.9	3.2
RBIS	0.7105	.163	4.35	0.55	0.11
DIFPOM	−0.2303	.012	−18.82	0.37	1.45
DIFPNR	−0.1300	.028	− 4.71	0.19	0.69
REGMATH	0.0426	.019	− 2.21	0.33	0.94
Regular Mathematical Item Type	*Multiple R = .69*		*R Squared = .48*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 355)*	*Mean*	*SD*
EQDEL	0.0548	.007	7.41	12.0	3.4
RBIS	0.8318	.210	3.96	0.56	0.11
DIFPOM	−0.2316	.014	−16.19	0.44	1.60
DIFPNR	−0.1254	.030	− 4.19	0.28	0.83
Quantitative Comparison Item Type	*Multiple R = .61*		*R Squared = .38*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 176)*	*Mean*	*SD*
EQDEL	0.0589	.010	6.07	11.7	2.9
RBIS	0.4770	.252	1.89	0.53	0.11
DIFPOM	−0.2264	.025	− 8.96	0.23	1.09
DIFPNR	—	—	—	0.0	0.0

Table 15. Correlations Based on SAT-Mathematical Data for Asian American Focal Group and White Reference Group

			Overall Mathematical Sections			
	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	*REGMATH*
MHD	1.00					
EQDEL	.22	1.00				
RBIS	.17	− .16	1.00			
DIFPOM	− .59	.05	− .08	1.00		
DIFPNR	− .01	.34	− .09	− .06	1.00	
REGMATH	.04	.04	.15	.07	.19	1.00
			Regular Mathematical Item Type			
	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	
MHD	1.00					
EQDEL	.22	1.00				
RBIS	.18	− .15	1.00			
DIFPOM	− .62	.00	− .07	1.00		
DIFPNR	− .02	.40	− .15	− .08	1.00	
			Quantitative Comparison Item Type			
	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	
MHD	1.00					
EQDEL	.24	1.00				
RBIS	.13	− .19	1.00			
DIFPOM	− .49	.19	− .16	1.00		
DIFPNR	—	—	—	—	—	

Table 16. Regression Results for Predicting MHD from SAT-Mathematical Item Characteristics (Females and Males)

Overall Mathematical Sections	*Multiple R = .61*		*R Squared = .38*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 534)*	*Mean*	*SD*
EQDEL	0.0073	.006	1.21	12.2	3.1
RBIS	−0.1800	.154	− 1.17	0.56	0.12
DIFPOM	−0.1906	.011	−16.99	0.46	1.58
DIFPNR	−0.0801	.026	− 3.10	0.10	0.67
REGMATH	0.0238	.018	1.31	0.33	0.94
Regular Mathematical Item Type	*Multiple R = .65*		*R Squared = .42*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 355)*	*Mean*	*SD*
EQDEL	0.0147	.007	2.06	12.3	3.3
RBIS	−0.0719	.191	− 0.38	0.57	0.11
DIFPOM	−0.1958	.013	−15.43	0.45	1.73
DIFPNR	−0.0853	.027	− 3.22	0.15	0.82
Quantitative Comparison Item Type	*Multiple R = .51*		*R Squared = .26*		
Independent Variable	*Weight*	*SE of Weight*	*t (df = 176)*	*Mean*	*SD*
EQDEL	−0.0130	.011	− 1.16	12.0	2.8
RBIS	−0.3903	.264	− 1.48	0.53	0.11
DIFPOM	−0.1722	.025	− 6.92	0.47	1.25
DIFPNR	—	—	—	0.0	0.0

Table 17. Correlations Based on SAT-Mathematical Data for Female Focal Group and Male Reference Group

Overall Mathematical Sections

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*	*REGMATH*
MHD	1.00					
EQDEL	− .16	1.00				
RBIS	.04	− .26	1.00			
DIFPOM	− .60	.32	− .13	1.00		
DIFPNR	− .13	.22	− .07	.06	1.00	
REGMATH	.03	.05	.15	− .01	.11	1.00

Regular Mathematical Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	− .13	1.00			
RBIS	.02	− .26	1.00		
DIFPOM	− .63	.30	− .07	1.00	
DIFPNR	− .15	. .26	− .11	.07	1.00

Quantitative Comparison Item Type

	MHD	*EQDEL*	*RBIS*	*DIFPOM*	*DIFPNR*
MHD	1.00				
EQDEL	− .25	1.00			
RBIS	.07	− .30	1.00		
DIFPOM	− .50	.39	− .30	1.00	
DIFPNR	—	—	—	—	—